

## O.R. Applications

# Recovering risk-neutral probability density functions from options prices using cubic splines and ensuring nonnegativity

Ana Margarida Monteiro <sup>a</sup>, Reha H. Tütüncü <sup>b,1</sup>, Luís N. Vicente <sup>c,\*,2</sup><sup>a</sup> *Faculdade de Economia, Universidade de Coimbra, Av. Dias da Silva, 165, 3004-512 Coimbra, Portugal*<sup>b</sup> *Department of Mathematical Sciences, 6113 Wean Hall, Carnegie Mellon University, Pittsburgh, PA 15213, USA*<sup>c</sup> *Departamento de Matemática, Universidade de Coimbra, 3001-454 Coimbra, Portugal*

Received 1 December 2005; accepted 22 February 2007

Available online 28 March 2007

## Abstract

We present a new approach to estimate the risk-neutral probability density function (pdf) of the future prices of an underlying asset from the prices of options written on the asset. The estimation is carried out in the space of cubic spline functions, yielding appropriate smoothness. The resulting optimization problem, used to invert the data and determine the corresponding density function, is a convex quadratic or semidefinite programming problem, depending on the formulation. Both of these problems can be efficiently solved by numerical optimization software.

In the quadratic programming formulation the positivity of the risk-neutral pdf is heuristically handled by posing linear inequality constraints at the spline nodes. In the other approach, this property of the risk-neutral pdf is rigorously ensured by using a semidefinite programming characterization of nonnegativity for polynomial functions.

We tested our approach using data simulated from Black–Scholes option prices and using market data for options on the S&P 500 Index. The numerical results we present show the effectiveness of our methodology for estimating the risk-neutral probability density function.

© 2007 Elsevier B.V. All rights reserved.

**Keywords:** Option pricing; Risk-neutral density estimation; Cubic splines; Quadratic programming; Semidefinite programming

## 1. Introduction

The risk-neutral probability measure is a fundamental concept in arbitrage pricing theory. By definition, a risk-neutral probability measure (RNPM) is a measure under which the current price of each security in the

\* Corresponding author.

E-mail addresses: [amonteiro@fe.uc.pt](mailto:amonteiro@fe.uc.pt) (A.M. Monteiro), [reha@cmu.edu](mailto:reha@cmu.edu) (R.H. Tütüncü), [lnv@mat.uc.pt](mailto:lnv@mat.uc.pt) (L.N. Vicente).

<sup>1</sup> Support for this author was provided by the National Science Foundation under Grants CCR-9875559 and DMS-0139911.

<sup>2</sup> Support for this author was provided by Centro de Matemática da Universidade de Coimbra, by FCT under Grants POCTI/35059/MAT/2000 and POCI/59442/2004, by the European Union under Grant IST-2000-26063, and by Fundação Calouste Gulbenkian. The author also thank the IBM T.J. Watson Research Center and the Institute for Mathematics and Its Applications for their local support.

economy is equal to the present value of the discounted expected value of its future payoffs given a risk-free interest rate. Fundamental theorems of asset pricing indicate that RNPMs are guaranteed to exist under an assumption of no arbitrage.

If a unique RNPM on the space of future states of an economy is given, we can price any security for which we can determine the future payoffs for each state in the state space. Therefore, a fundamental problem in asset pricing is the identification of a risk-neutral probability measure. While the dynamics of an economy and the parameters for its stochastic models are not directly observable, one can infer some information about these dynamics from the current prices of the securities in this economy. In particular, one can extract one or more implied risk-neutral densities of the future price of a security that are consistent with the prices of options written on that security. When there are multiple RNPMs consistent with the observed prices, one may try to choose the “best” one, according to some criterion. We address this problem in this article using optimization models.

For a stock or index, the set of possible future states can be represented as an interval or ray, discretized if appropriate or necessary. In most cases, the number of states in this state space is much larger than the number of observed prices, resulting in a problem with many more variables than equations. This underdetermined problem has many potential solutions and we cannot obtain an unique or sensible solution without imposing some additional structure into the risk-neutral probability measure we are looking for.

The type of additional structure imposed has been the differentiating feature of the existing approaches to the problem of identifying implied RNPMs. These approaches can be broadly classified as parametric and nonparametric techniques and are reviewed by Jackwerth [18], and Bondarenko [9] (see also Section 2). Parametric methods choose a distribution family (or a mixture of distributions) and then try to identify the parameters for these distributions that are consistent with the observed prices [6,22]. In nonparametric techniques, one achieves more flexibility by allowing general functional forms. Structure is introduced either using prior distributions or smoothness restrictions. Our approach fits into this last category and we ensure the desired smoothness of the RNPM using spline functions.

Spline functions are piecewise polynomial functions that assume a predetermined value at certain points (*knots*) and satisfy certain smoothness properties. Other authors have also used spline fitting techniques in the context of risk-neutral density estimation, see [3,13]. In contrast to existing techniques, we allow the displacement of spline knots in a superset of the set of points corresponding to option strikes. The additional set of knots makes our model flexible and we use this flexibility to optimize the fit of the spline function to the observed prices. The basic formulation, without requiring the nonnegativity of the risk-neutral probability density function (pdf), is a convex quadratic programming (QP) problem.

Two strategies to impose the nonnegativity of the RNPM are presented and discussed in this paper. The first and simpler strategy is to require the estimated pdf to remain nonnegative at the spline nodes. This scheme maintains the QP structure of the problem since it brings only linear inequality constraints to the basic formulation. However, there is no guarantee of nonnegativity between the spline nodes. Our second approach replaces the basic QP formulation with a semidefinite programming (SDP) formulation but rigorously ensures the nonnegativity of the estimated pdf in its entire domain. It is based on an SDP characterization of nonnegative polynomial functions due to Bertsimas and Popescu [5] and requires additional variables and linear equality constraints as well as semidefiniteness constraints on some matrix variables. To our knowledge, this is the first spline function approach to risk-neutral density estimation with a positivity guarantee.

The rest of this paper is organized as follows: In Section 2, we provide the definition of RNPMs and briefly discuss some of the existing approaches. In Section 3, we discuss our spline approximation approach to RNPMs and develop our basic QP optimization model. The treatment of nonnegativity is given in Section 4. Section 5 is devoted to a numerical study of our approach both with simulated and market data. We provide a brief conclusion in Section 6.

## 2. Risk-neutral probability measures

We consider the following one-period economy: There are  $n$  securities whose current prices are given by  $s_0^i$  for  $i = 1, \dots, n$ . At the end of the current period, the economy will be in one of the states from the state space  $\Omega$ . If the economy reaches state  $\omega \in \Omega$  at the end of the current period, security  $i$  will have the payoff  $s_1^i(\omega)$ . We

assume that we know all  $s_0^i$ 's and  $s_1^i(\omega)$ 's but do not know the particular terminal state  $\omega$ , which will be determined randomly.

As an example of the set-up explained in the previous paragraph, we consider a particular security (stock, index, etc.) and let the  $n$  securities be financial options written on this stock. Here,  $\Omega$  denotes the state space for the terminal price of the underlying stock and  $s_1^i(\omega)$  denotes the payoff of the option  $i$  when the underlying stock price is  $\omega$  at termination. For example, if option  $i$  is a European call with strike price  $K_i$  to be exercised at the end of the current period, we would have  $s_1^i(\omega) = (\omega - K_i)^+$ .

Next, we give a definition of RNPMs:

**Definition 1.** Consider the economy described above. Let  $r$  denote the risk-free interest rate in the period  $[t, T]$ . A risk-neutral probability measure in the

- *discrete case* and on the state space  $\Omega = \{\omega_1, \omega_2, \dots, \omega_m\}$  is defined by a vector of positive numbers  $p_1, p_2, \dots, p_m$  such that
  1.  $\sum_{j=1}^m p_j = 1$ ,
  2.  $s_0^i = e^{-r(T-t)} \sum_{j=1}^m p_j s_1^i(\omega_j)$ ,  $i = 1, \dots, n$ ;
- *continuous case* and on the state space  $\Omega = (a, b)$  is defined by a density function  $p : \mathbb{R} \rightarrow \mathbb{R}_+$  such that
  1.  $\int_a^b p(\omega) d\omega = 1$ ,
  2.  $s_0^i = e^{-r(T-t)} \int_a^b p(\omega) s_1^i(\omega) d\omega$ ,  $i = 1, \dots, n$ .

It is well known that the existence of a risk-neutral probability measure is strongly related to the absence of arbitrage opportunities as expressed in the *First Fundamental Theorem of Asset Pricing* (see [15]). We first give an informal definition of arbitrage and then state this theorem:

**Definition 2.** An *arbitrage* is a trading strategy

- that has a positive initial cash flow and has no risk of a loss later, or
- that requires no initial cash input, has no risk of a loss, and a positive probability of making profits in the future.

**Theorem 1.** A risk-neutral probability measure exists if and only if there are no arbitrage opportunities.

Definition 1 and Theorem 1 illustrate why we care about risk-neutral probability measures. When a RNPM is available on the future states of an economy, any security whose future payoffs in all these future states is known can be reliably priced. Furthermore, if we assume that arbitrage opportunities do not exist as commonly done in the financial literature, we know that an RNPM must exist. Then the question becomes, how do we find or estimate an RNPM on the future states of an economy?

One very important and potentially tractable instance of this problem arises when the economy consists of a security (say a stock or a market index) that we will call the underlying security and a number of derivative securities based on this underlying security. We are interested in the RNPM for the price of the underlying security at a fixed future date. As we argued in Section 1, since the payoffs of the derivatives depend on the future values of the underlying asset, we can use the prices of these derivatives to get information about the probability distribution of the future values of the underlying. We can say that the prices of option contracts contain some information about the market expectations, namely a possible correspondence between the price of the underlying and its strike.

The estimation of the risk-neutral densities serves several purposes. The most common application of risk-neutral densities is related to the setting we discussed in the previous paragraph. Given the observed prices of derivative securities on a particular underlying asset, one can use the RNPM derived from these prices to determine an arbitrage-free price (or limits on arbitrage-free prices) of new, exotic derivatives with complicated payoff structures on the same underlying security. By providing a mechanism for extrapolating information from observed derivative prices, risk-neutral densities can also be used to detect arbitrage opportunities. Another relevant application is related to capturing the *market sentiment*. A risk-neutral density can compactly summarize the collective expectation of market participants regarding the fluctuations of asset

prices in the future (see, e.g., [6]). This information may assume a greater importance under unusual economic or political scenarios (see, e.g., [22]). The risk-neutral densities are also used to determine new risk-adjusted densities that accommodate investors' risk preferences, namely the preferences that derive from the interaction between the market participants.

Mathematically, the problem of estimating risk-neutral densities provides an interesting challenge. From an optimizer's perspective, the problem is to find the best RNPM estimate given the available information. The price of each additional security provides an additional piece of information on the structure of the density and makes the set of densities consistent with all observed prices smaller. Structural properties of probability measures (nonnegativity, integrate to one) provide additional constraints that are not always straightforward to impose.

There are several approaches, reported in the literature, to derive risk-neutral probabilities from observed options prices (see the surveys in [6,9,18]). One can either assume that the density belongs to a parametric family of distributions and reduce the problem to a simple parameter-fitting problem, or use a nonparametric method. In the context of the Black–Scholes model [7], it is well known that the geometric Brownian motion followed by the underlying asset price implies a lognormal risk-neutral pdf. In practice, parametric methods derive the risk-neutral pdf's from a set of statistical distributions and the set of observational data. A specific parameterized form for the density function on the underlying asset (such as a mixture of lognormals; see [6,22]), or for the stochastic process [7,14] is assumed and then the optimal parameters are identified by fitting the data (see also [4,24]).

Nonparametric methods are more flexible as they essentially allow an infinite number of parameters. Among the nonparametric approaches, and following Bondarenko [9], we can point out: implied trees methods [23], kernel methods [2], maximum-entropy methods [12,26], methods applied to the volatility smile [3,11,13,25], methods applied to the risk-neutral probability [19,21]. Other methods share both parametric and nonparametric features [1,9]. Here, we follow the nonparametric approach.

Several of the methods mentioned above, including both parametric and nonparametric ones, suffer from a lack of guarantee of nonnegativity of the risk-neutral density they generate. Jackwerth [18] notes that parametric methods based on the expansion of a basic risk-neutral density can generate densities with negative values. Rubinstein [24] states that for some set of parameters the risk-neutral densities generated violate the nonnegativity constraint. The methods proposed by Shimko [25] and developed by Aparicio and Hodges [4], Bliss and Panigirtzoglou [8], and Campa et al. [13] can produce negative probabilities. Jackwerth [18] as well as Brunner and Hafner [11] mention that, for these methods, the nonnegativity of the density must be checked since there is no guarantee of nonnegativity. Bondarenko [9] corroborates the latter and indicates that also [19] may obtain negative probabilities. These concerns lead us to search for a guarantee of nonnegativity for the risk-neutral density. As we will later see, we show that it is possible to guarantee the nonnegativity of the risk-neutral density in a rigorous manner by reformulating the problem using tools from semidefinite programming.

### 3. The basic formulation using splines

As discussed in Section 1, one of the desired structural properties of a RNPM estimate is smoothness. The strategy developed in this section guarantees appropriate smoothness of the risk-neutral pdf by estimating it using cubic splines (see the Appendix of this paper). The estimation is carried out by the solution of an optimization problem where the optimization variables are the parameters of the spline functions.

#### 3.1. The quadratic programming formulation

We now formulate an optimization problem with the objective of finding a risk-neutral pdf described by cubic splines for future values of an underlying security that provides a best fit with the observed option prices on this security.

For the security under consideration, we fix an exercise date, a range  $[a, b]$  for possible terminal values of the price of the underlying security at the exercise date of the options, and an interest rate  $r$  for the period

between now and the exercise date. The other inputs to our optimization problem are the chosen expiration date and market prices,  $C_K$  for call options and  $P_K$  for put options, on the chosen underlying security, with strike price  $K$ . Let  $\mathcal{C}$  and  $\mathcal{P}$ , respectively, denote the set of strike prices  $K$  for which reliable market prices  $C_K$  and  $P_K$  are available. For example,  $\mathcal{C}$  may denote the strike prices of call options that were traded on the day that the problem is formulated.

Next, we consider a super-structure for the spline approximation to the risk-neutral pdf, meaning that we choose how many knots to use, where to place the knots and what kind of polynomial (quadratic, cubic, etc.) functions to use. For example, one may decide to use cubic splines (as we do in this paper) and  $n_s + 1$  equally spaced knots. The parameters of the polynomial functions that comprise the spline function will be the variables of the optimization problem we are formulating. For cubic splines with  $n_s + 1$  knots, we will have  $4n_s$  variables  $(\alpha_s, \beta_s, \gamma_s, \delta_s)$ , for  $s = 1, \dots, n_s$ . Collectively, we will represent these variables by  $y \in \mathbb{R}^{4n_s}$ . For all  $y$  chosen so that the corresponding polynomial functions  $f_s$  satisfy the systems (20)–(23) of the Appendix, we will have a particular (natural) spline function defined on the interval  $[a, b]$ . Let  $p_y(\omega)$  denote this function. Note that we do not impose the constraints given in (19) because the values of the pdf we are approximating are unknown and will be the result of the solution of an optimization problem.

By imposing the following additional restrictions we make sure that  $p_y$  is a probability density function:

$$p_y(\omega) \geq 0, \quad \forall \omega \in [a, b], \quad (1)$$

$$\int_a^b p_y(\omega) d\omega = 1. \quad (2)$$

In practice the requirement (2) is easily imposed by including the following constraint in the optimization problem:

$$\sum_{s=1}^{n_s} \int_{x_s}^{x_{s+1}} f_s(\omega) d\omega = 1. \quad (3)$$

One can easily see that this is a linear constraint in the components  $(\alpha_s, \beta_s, \gamma_s, \delta_s)$  of the optimization variable  $y$ . The treatment of (1) is postponed to the next section and is ignored until the end of this section.

Next, we define the discounted expected value of the terminal value of each option using  $p_y$  as the risk-neutral probability density function (see [10,14]):

$$C_K(y) = e^{-r(T-t)} \int_a^b p_y(\omega) (\omega - K)^+ d\omega, \quad (4)$$

$$P_K(y) = e^{-r(T-t)} \int_a^b p_y(\omega) (K - \omega)^+ d\omega. \quad (5)$$

If  $p_y$  was the actual risk-neutral probability density function, the quantities  $C_K(y)$  and  $P_K(y)$  would be the fair values of the call and put options with strikes  $K$ . The quantity

$$(C_K - C_K(y))^2,$$

measures the squared difference between the observed value and discounted expected value considering  $p_y$  as the risk-neutral pdf. Now consider the overall residual least squares function for a given  $y$ :

$$E(y) = \sum_{K \in \mathcal{C}} (C_K - C_K(y))^2 + \sum_{K \in \mathcal{P}} (P_K - P_K(y))^2. \quad (6)$$

The objective now is to choose  $y$  such that  $E(y)$  is minimized subject to the constraints already mentioned. The resulting optimization problem is a convex quadratic programming problem corresponding to the following formulation:

$$\min_y E(y) \quad \text{s.t. (20)–(23), (3).} \quad (7)$$

### 3.2. Functions $C_K(y)$ and $P_K(y)$

We now look at the structure of problem (7) in more detail. In particular, we evaluate the function  $C_K(y)$ . Consider a call option with strike  $K$  such that  $x_\ell \leq K < x_{\ell+1}$ . Recall that  $y$  denotes a collection of variables  $(\alpha_s, \beta_s, \gamma_s, \delta_s)$  for  $s = 1, \dots, n_s$  and that  $x_1 = a, x_2, \dots, x_{n_s}, x_{n_s+1} = b$  represent the locations of the knots. The formula for  $C_K(y)$  can be derived as follows:

$$\begin{aligned} e^{r(T-t)} C_K(y) &= \int_a^b p_y(\omega)(\omega - K)^+ d\omega = \sum_{s=\ell}^{n_s} \int_{x_s}^{x_{s+1}} p_y(\omega)(\omega - K)^+ d\omega \\ &= \int_K^{x_{\ell+1}} p_y(\omega)(\omega - K) d\omega + \sum_{s=\ell+1}^{n_s} \int_{x_s}^{x_{s+1}} p_y(\omega)(\omega - K) d\omega \\ &= \int_K^{x_{\ell+1}} (\alpha_\ell \omega^3 + \beta_\ell \omega^2 + \gamma_\ell \omega + \delta_\ell)(\omega - K) d\omega + \sum_{s=\ell+1}^{n_s} \int_{x_s}^{x_{s+1}} (\alpha_s \omega^3 + \beta_s \omega^2 + \gamma_s \omega + \delta_s)(\omega - K) d\omega. \end{aligned}$$

One can easily see that this expression for  $C_K(y)$  is linear in the components  $(\alpha_s, \beta_s, \gamma_s, \delta_s)$  of the optimization variable  $y$ . A similar formula can be derived for  $P_K(y)$ . Another relevant aspect that should be pointed out is that the formula for  $C_K(y)$  will involve coefficients of the type  $x_s^5$  which can, and in fact do, make the Hessian matrix of the QP problem (7) severely ill-conditioned.

### 4. Guaranteeing nonnegativity

The simplest way to deal with the requirement of nonnegativity of the risk-neutral pdf is to weaken condition (1), requiring the cubic spline approximation to be nonnegative only at the knots:

$$f_s(x_s) \geq 0, \quad s = 1, \dots, n_s \quad \text{and} \quad f_{n_s}(x_{n_s+1}) \geq 0. \quad (8)$$

Then, the basic QP formulation changes to

$$\min_y E(y) \quad \text{s.t.} \quad (20)–(23), \quad (3), \quad (8). \quad (9)$$

One can easily see that problem (9) is still a convex quadratic programming problem, since (8) are linear inequalities in the optimization variables. The drawback of this strategy is the lack of guarantee of nonnegativity of the spline functions between the spline knots. This heuristic strategy proved sufficient to obtain non-negative pdf estimates in most of our experiments some of which are reported in Section 5. However, in some instances pdf estimates assumed negative values between knots. Since our aim is to estimate a probability density function, estimates with negative values are not acceptable.

In what follows, we develop an alternative optimization model where the nonnegativity of the resulting risk-neutral pdf estimate is rigorously guaranteed. The cost we must pay for this guarantee is an increase in the complexity of the optimization problem. Indeed, the new model involves semidefiniteness restrictions on some matrices related to new optimization variables. While the resulting problem is still a convex optimization problem and can be solved with standard conic and semidefinite optimization software (see, e.g., [27]), it is also more expensive to solve than a convex QP.

The model we consider is based on necessary and sufficient conditions for the nonnegativity of a single variable polynomial in intervals, as well as on rays and on the whole real line. This characterization is due to Bertsimas and Popescu [5] and is stated in the next proposition.

**Proposition 1** (Proposition 1(d), [5]). *The polynomial  $g(x) = \sum_{r=0}^k y_r x^r$  satisfies  $g(x) \geq 0$  for all  $x \in [a, b]$  if and only if there exists a positive semidefinite matrix  $X = [x_{ij}]_{i,j=0,\dots,k}$  such that*

$$\sum_{i,j:i+j=2\ell-1} x_{ij} = 0, \quad \ell = 1, \dots, k, \quad (10)$$

$$\sum_{i,j:i+j=2\ell} x_{ij} = \sum_{m=0}^{\ell} \sum_{r=m}^{k+m-\ell} y_r \binom{r}{m} \binom{k-r}{\ell-m} a^{r-m} b^m, \quad (11)$$



$$\ell = 0, \dots, k, \quad (12)$$

$$X \succeq 0. \quad (13)$$

In the statement of the proposition above, the notation  $\binom{r}{m}$  stands for  $\frac{r!}{m!(r-m)!}$  and  $X \succeq 0$  indicates that the matrix  $X$  is symmetric and positive semidefinite. For the cubic polynomial  $f_s(x) = \alpha_s x^3 + \beta_s x^2 + \gamma_s x + \delta_s$  we have the following corollary:

**Corollary 1.** *The polynomial  $f_s(x) = \alpha_s x^3 + \beta_s x^2 + \gamma_s x + \delta_s$  satisfies  $f_s(x) \geq 0$  for all  $x \in [x_s, x_{s+1}]$  if and only if there exists a  $4 \times 4$  matrix  $X^s = [x_{ij}^s]_{i,j=0,\dots,3}$  such that*

$$\begin{aligned} x_{ij}^s &= 0, \quad \text{if } i+j \text{ is } 1 \text{ or } 5, \\ x_{03}^s + x_{12}^s + x_{21}^s + x_{30}^s &= 0, \\ x_{00}^s &= \alpha_s x_s^3 + \beta_s x_s^2 + \gamma_s x_s + \delta_s, \\ x_{02}^s + x_{11}^s + x_{20}^s &= 3\alpha_s x_s^2 x_{s+1} + \beta_s (2x_s x_{s+1} + x_s^2) + \gamma_s (x_{s+1} + 2x_s) + 3\delta_s, \\ x_{13}^s + x_{22}^s + x_{31}^s &= 3\alpha_s x_s x_{s+1}^2 + \beta_s (2x_s x_{s+1} + x_{s+1}^2) + \gamma_s (x_s + 2x_{s+1}) + 3\delta_s, \\ x_{33}^s &= \alpha_s x_{s+1}^3 + \beta_s x_{s+1}^2 + \gamma_s x_{s+1} + \delta_s, \\ X^s &\succeq 0. \end{aligned} \quad (14)$$

Observe that the positive semidefiniteness of the matrix  $X^s$  implies that the first diagonal entry  $x_{00}^s$  is nonnegative, which corresponds to our earlier requirement  $f_s(x_s) \geq 0$ . In light of Corollary 1, we see that introducing the additional variables  $X^s$  and the constraints (14), for  $s = 1, \dots, n_s$ , into the earlier quadratic programming problem (7), we obtain a new optimization problem which necessarily leads to a risk-neutral pdf that is non-negative in its entire domain. The new formulation has the following form:

$$\min_{y, X^1, \dots, X^{n_s}} E(y) \quad \text{s.t. (20)–(23), (3), [(14), } s = 1, \dots, n_s]. \quad (15)$$

All constraints in (15), with the exception of the positive semidefiniteness constraints  $X^s \succeq 0$ ,  $s = 1, \dots, n_s$ , are linear in the optimization variables  $(\alpha_s, \beta_s, \gamma_s, \delta_s)$  and  $X^s$ ,  $s = 1, \dots, n_s$ . The positive semidefiniteness constraints are convex constraints and thus the resulting problem can be reformulated as a (convex) semidefinite programming problem with a quadratic objective function.

For appropriate choices of the vectors  $c, f_i, g_k^s$ , and matrices  $Q$  and  $H_k^s$ , we can rewrite problem (15) in the following equivalent form:

$$\begin{aligned} \min_{y, X^1, \dots, X^{n_s}} \quad & c^\top y + \frac{1}{2} y^\top Q y \\ \text{s.t.} \quad & f_i^\top y = b_i, \quad i = 1, \dots, 3n_s, \\ & H_k^s \bullet X^s = 0, \quad k = 1, 2, \quad s = 1, \dots, n_s, \\ & (g_k^s)^\top y + H_k^s \bullet X^s = 0, \quad k = 3, 4, 5, 6, \quad s = 1, \dots, n_s, \\ & X^s \succeq 0, \quad s = 1, \dots, n_s, \end{aligned} \quad (16)$$

where  $\bullet$  denotes the trace matrix inner product.

We should note that standard semidefinite optimization software such as SDPT3 [27] can solve only problems with linear objective functions. Since the objective function of (16) is quadratic in  $y$  a reformulation is necessary to solve this problem using SDPT3 or other SDP solvers. We replace the objective function with  $\min t$  where  $t$  is a new artificial variable and impose the constraint  $t \geq c^\top y + \frac{1}{2} y^\top Q y$ . This new constraint can be expressed as a second-order cone constraint after a simple change of variables; see, e.g., [20]. This final formulation is a standard form *conic optimization* problem, which is a class of problems that contains

semidefinite programming and second-order cone programming as special classes. Since SDPT3 can solve standard form conic optimization problems we used this formulation in our numerical experiments.

## 5. Numerical experiments

In this section, we report some numerical experiments obtained with the methodologies introduced in this paper to estimate the risk-neutral pdf, namely the approaches that led to the formulation of problems (9) and (15). We also consider a mixture of lognormals model and compare the three approaches.

We have applied the active set method provided by MATLAB to solve the convex QP problem (9) and the MATLAB-based interior-point code SDPT3 [27] to solve the SDP problem (15) (more precisely its reformulation described at the end of the last section). The performance of these two approaches is illustrated with four different data sets, one generated from a Black–Scholes model and the other three extracted from the S&P 500 Index.

Numerically, we solved scaled versions of both the QP problem (9) and the SDP problem (16). The need for scaling the data of these problems results from the fact that the Hessian matrix in (7), which appears in both problems, is highly ill-conditioned, as we have already pointed out in Section 3.2. Since the magnitude of  $\omega$  plays a relevant role in the size of the entries of this Hessian matrix, we used as our reference scaling factor the average value of the components of the vector of the knots. Let us call this average value  $x_{\text{avg}}$ . Then each knot  $x_s$ ,  $s = 1, \dots, n_s + 1$ , is scaled by  $x_{\text{avg}}$  and replaced by  $x'_s = x_s/x_{\text{avg}}$ . Such a scaling amounts at the end to scale the variables  $\alpha_s, \beta_s, \gamma_s, \delta_s$  corresponding to the spline coefficients by, respectively,  $a, b, c, d$ , whose values depend on  $x_{\text{avg}}$  as well as on the expressions for the integrations given in Section 3.2. The problem is then solved in the scaled variables  $\alpha'_s, \beta'_s, \gamma'_s, \delta'_s$ ,  $s = 1, \dots, n_s$ . We also multiply each term of the objective function in (7) by  $1/x_{\text{avg}}^2$ . The unscaled solution is recovered by the formulas  $(\alpha_s, \beta_s, \gamma_s, \delta_s) = (a\alpha'_s, b\beta'_s, c\gamma'_s, d\delta'_s)$ ,  $s = 1, \dots, n_s$ .

### 5.1. Mixture of lognormals approach

We compared our QP and SDP approaches with an approach based on mixture of lognormals. Following the procedures presented by Bahra [6], we considered the minimization of

$$E(z) = \sum_{K \in \mathcal{C}} (C_K - C_K(z))^2 + \sum_{K \in \mathcal{P}} (P_K - P_K(z))^2, \quad (17)$$

where

$$C_K(z) = e^{-r(T-t)} \int_a^b p_z(\omega)(\omega - K)^+ d\omega,$$

$$P_K(z) = e^{-r(T-t)} \int_a^b p_z(\omega)(K - \omega)^+ d\omega.$$

Here,  $z$  is a vector of five parameters  $(\gamma, \alpha_1, \alpha_2, \beta_1, \beta_2)$  storing the means and standard deviations of two lognormals and the weight of their linear combination. In fact, the pdf in this approach takes the form

$$p_z(\omega) = \gamma \log n(\omega; \alpha_1, \beta_1) + (1 - \gamma) \log n(\omega; \alpha_2, \beta_2),$$

$$\log n(\omega; \alpha, \beta) = \frac{1}{\omega \sqrt{2\pi\beta}} e^{-\frac{(\ln(\omega) - \alpha)^2}{2\beta^2}},$$

$$\beta = \sigma \sqrt{T - t} \quad \text{and} \quad \alpha = \ln(S_t) + \mu(T - t) - \frac{1}{2}\beta^2,$$

where  $\mu$  is the expected return of the underlying asset and  $\sigma$  its volatility. The parameters of the lognormals are obtained by solving the constrained nonlinear least squares problem (LN2) defined by the objective function (17) and one constraint

$$\min_z E(z) \quad \text{s.t.} \quad \gamma e^{\alpha_1 + \frac{1}{2}\beta_1^2} + (1 - \gamma) e^{\alpha_2 + \frac{1}{2}\beta_2^2} = S_t e^{r(T-t)}. \quad (18)$$



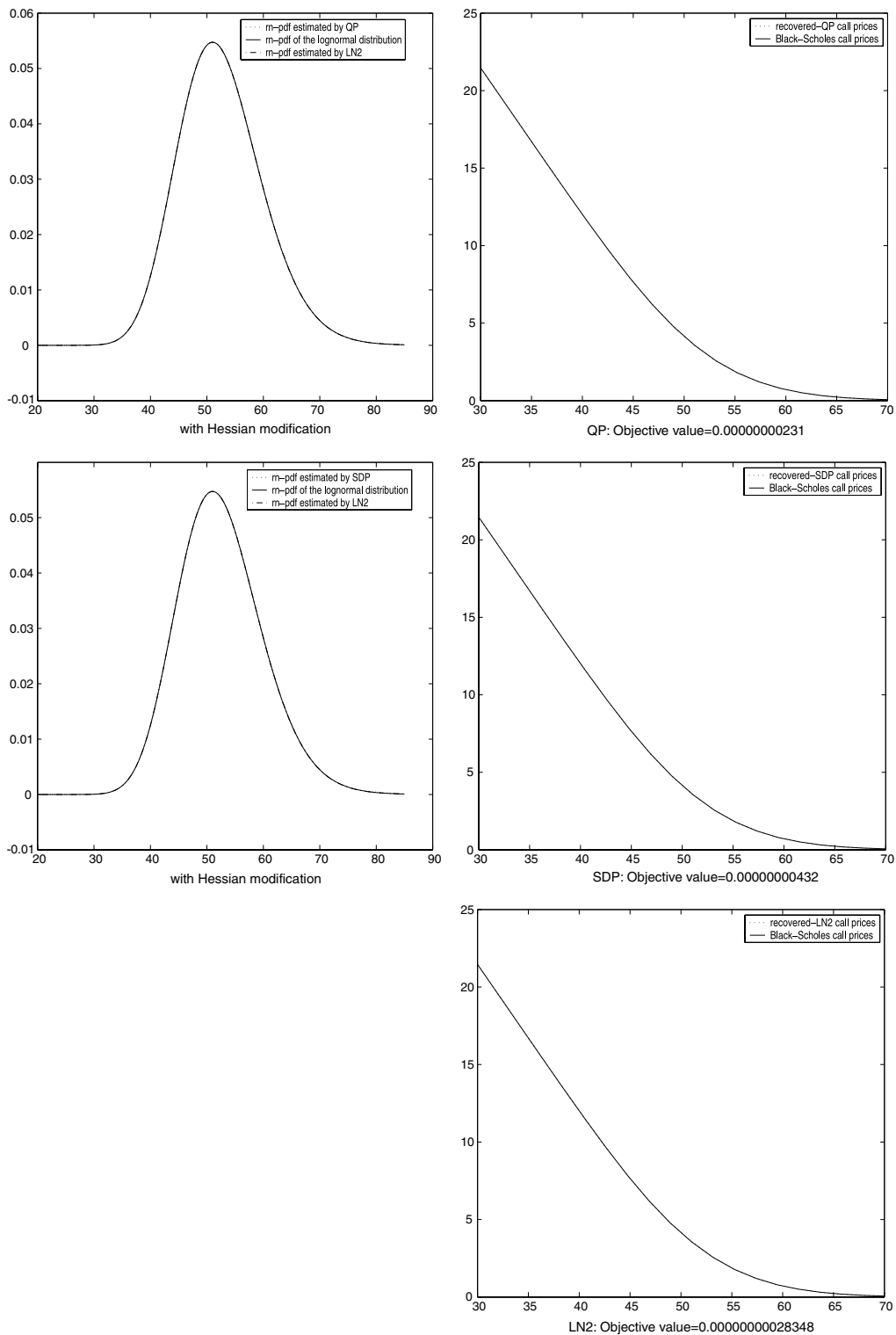


Fig. 1. Recovered probability density functions from data generated by a Black-Scholes model using QP, SDP, and LN2 approaches (left plots). Fitted recovered expected prices for these approaches (right plots).

The constraint is equivalent to

$$\gamma S_t e^{\mu_1(T-t)} + (1 - \gamma) S_t e^{\mu_2(T-t)} = S_t e^{r(T-t)}$$

and it is basically saying that in the absence of arbitrage the estimated forward price of the underlying security must coincide with its forward price [6].

Our numerical implementation was based on MATLAB's optimization solver FMINCON. We took into consideration the fact that  $E(z)$  is nonlinear and nonconvex by scaling the problem appropriately and by paying special attention to the initial point provided to FMINCON (in the Black–Scholes data set, for instance, we used the exact values of the targeted lognormal as initial guesses).

## 5.2. Black–Scholes data

Our first test was conducted with simulated option prices. We generated Black–Scholes options prices using the function BLSPRICE provided by the Financial Toolbox of MATLAB. This function computes the value of the call or put option in agreement with the Black–Scholes formula. To generate the data we must supply the current value of the underlying asset, the exercise price, the risk-free interest rate, the time to maturity of the option, the volatility, and the dividend rate.

The call and put option prices were generated considering 50 as the current price for the underlying asset, 0.1 as the risk-free interest rate, a time to maturity of 0.5, a volatility of 0.2, and no dividend rate. We considered 20 call options with (equally spaced) strikes varying from 30 to 70. The number of knots was set to 25 and the knots were equally spaced between 20 and 85. The risk-neutral pdf corresponding to the prices generated from these data is known to be the following lognormal density function:

$$p(\omega) = \text{logn}(\omega; \alpha, \beta),$$

where  $\mu = 0.1$ ,  $\sigma = 0.2$ ,  $T - t = 0.5$ , and  $S_t = 50$ . This function is depicted in solid lines in the plots of Fig. 1.

We solved the scaled instances of problems (9) and (16) defined by the Black–Scholes data and the scaling reported above. We also solved problem (18) for the same data set. The plots of the recovered probability density functions are depicted in Fig. 1.

In our formulations, the Hessian matrix is theoretically known to be positive semidefinite. However, it is also highly rank-deficient and, due to round-off errors, it exhibits small negative eigenvalues, around  $-10^{-19}$ . These negative eigenvalues proved to be troublesome for MATLAB's active set QP. The scaling reduced significantly the ill-conditioning of this matrix, allowing a relatively accurate eigenvalue computation. We have modified the Hessian matrix, by adding a multiple  $\xi$  of the identity to the scaled Hessian matrix, using as coefficient  $\xi = |\lambda_{\min}|$ . Under this adjustment, the modified scaled Hessian becomes numerically positive definite. In the SDP approach, this value for  $\xi$  does not guarantee a numerically positive semidefinite matrix, and had to be increased to  $\xi = |\lambda_{\min}|10^3$ .

It can be seen from the left plots of Fig. 1 that the pdf computed adjusted perfectly well to the lognormal one, for all the approaches (QP, SDP, and LN2). We also point out that the expected prices of the call options computed using the recovered risk-neutral pdf adjusted very well to the Black–Scholes prices (see right plots of Fig. 1). The residuals for these adjustments, computed according to (6) in the QP and SDP cases and according to (17) in the LN2 approach, are reported in Table 1.

The order of approximation is almost the same for all approaches. It should be nevertheless stressed how well the recovered pdf obtained by QP and SDP exhibited the desired lognormality property.

In order to guarantee numerical success of the QP and SDP approaches, the initial and final spline knots must be chosen within a minimum distance from the first and last strikes. We found that this distance is around 6% of the amplitude of the vector of strikes. We chose the number of knots not much bigger than

Table 1  
Residuals for the price option adjustments (Black–Scholes data)

	LN2	QP	SDP
BS data set	$2.83 \times 10^{-10}$	$2.31 \times 10^{-9}$	$4.32 \times 10^{-9}$

the number of strikes, but the pdf shapes exhibit some robustness when we consider a different number of knots. In fact, for this Black–Scholes data set, we also obtain a good estimative for the pdf if we consider +40, +35, +20, or +15 knots. When we consider a logarithmically spaced vector of knots instead of a linearly spaced one (corresponding to the results reported above), the results remain the same: there is no difference in the shape of the pdf and the residuals are still very low ( $2.83 \times 10^{-10}$  for LN2,  $2.07 \times 10^{-9}$  for QP, and  $1.37 \times 10^{-7}$  for SDP).

### 5.3. S&P 500 data

The second set of tests were conducted with publicly available market price data. We collected information related to European call and put options on S&P 500 Index traded in the Chicago Board of Options Exchange (CBOE) on April 29, 2003 with maturity on May 17 (data set 1), on March 24, 2004 with maturity on April 17 (data set 2), and on March 24, 2004 with maturity on June 20 (data set 3). We chose this market because it is one of the most dynamic and liquid options markets in the world.

The interest rate was obtained from the Federal Reserve Bank of New York. We considered a Treasury Bill with time to expiration as close as possible to the time of expiration of the options.

#### 5.3.1. Preprocessing the data

As indicated in Section 2, a risk-neutral probability measure exists if and only if there are no arbitrage opportunities. It is possible, however, to observe arbitrage opportunities in the prices of illiquid derivative securities. These prices do not reflect true arbitrage opportunities – once these securities start trading, their prices will be corrected and arbitrage will not be realized.

Still, in order to have meaningful solutions for the optimization problems that we formulated in the previous sections, it is necessary to use prices in these optimization models which contain no arbitrage opportunities. Thus, before solving these problems we need to eliminate prices with arbitrage violations such as absence of monotonicity. The following theorem establishes necessary and sufficient conditions for the absence of arbitrage in the prices of European call options with concurrent expiration dates:

**Theorem 2** (Herzel [17]). *Let  $K_1 < K_2 < \dots < K_n$  denote the strike prices of European call options written on the same underlying security with the same maturity, and let  $C_i$  denote the current prices of these options. These securities do not contain any arbitrage opportunities if and only if the prices  $C_i$  satisfy the following conditions:*

1.  $C_i > 0$ ,  $i = 1, \dots, n$ .
2.  $C_i > C_{i+1}$ ,  $i = 1, \dots, n - 1$ .
3. The piecewise linear function  $C(K)$  with break-points at  $K_i$ 's and satisfying  $C(K_i) = C_i$ ,  $i = 1, \dots, n$ , is strictly convex in  $[K_1, K_n]$ .

Theorem 2 provides us with a simple mechanism to eliminate “artificial” arbitrage opportunities from the prices we use. In our numerical experiments, after gathering price data for call and put options from the S&P 500 Index, we first eliminated options whose prices were outside the bid-ask interval, and then we generated call option prices from each one of the put option prices using the put-call parity. In cases where there was already a call option with a matching strike price, in the event that the price of the traded call option did not coincide with the price obtained from the put option price using put-call parity, we used the price corresponding to the option with the higher trading volume. After obtaining a fairly large set of call option prices in this manner, we tested for monotonicity and strict convexity in these call prices as indicated by Theorem 2. After the prices that violated these conditions had been removed, we formulated and solved the optimization problems, as outlined in Section 4 for QP and SDP and in Section 5.1 for LN2.

In order to guarantee the quality of the data we collected another piece of information related to the market options: the trading volume (see [16]). It is known that end-of-day settlement prices can contain options that are not very liquid and these prices may not reflect the true market prices. Inaccurate prices are usually related to thinly traded options. In contrast, options with higher volume represent better the “market sentiment” and the investors expectations. We incorporated the trading volume in our problem formulation by modifying the objective function of problems (9) and (16) in the following way:

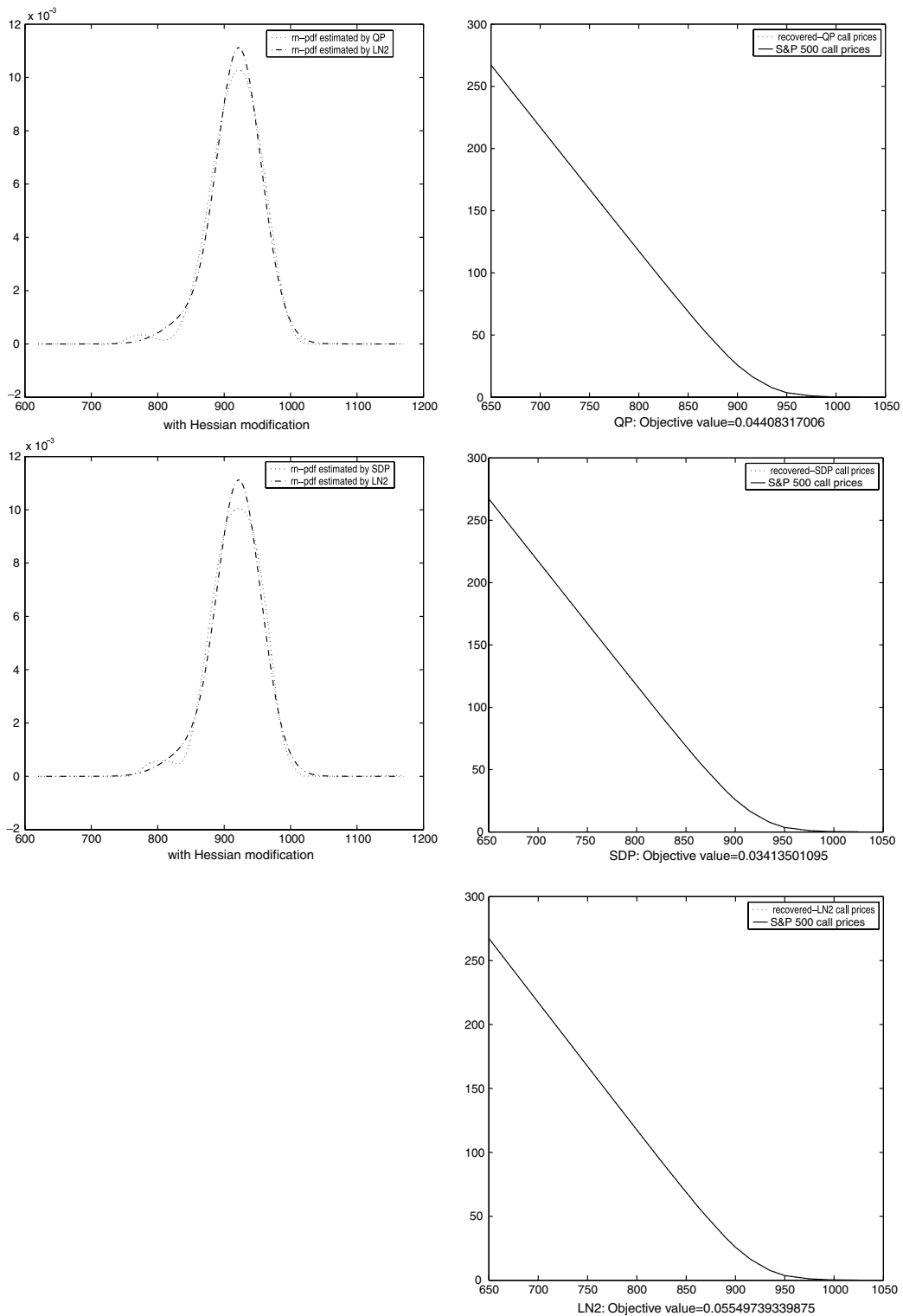


Fig. 2. Recovered probability density functions from S&P 500 Index data using QP, SDP, and LN2 approaches (left plots). Fitted recovered expected prices for these approaches (right plots). Data set 1.

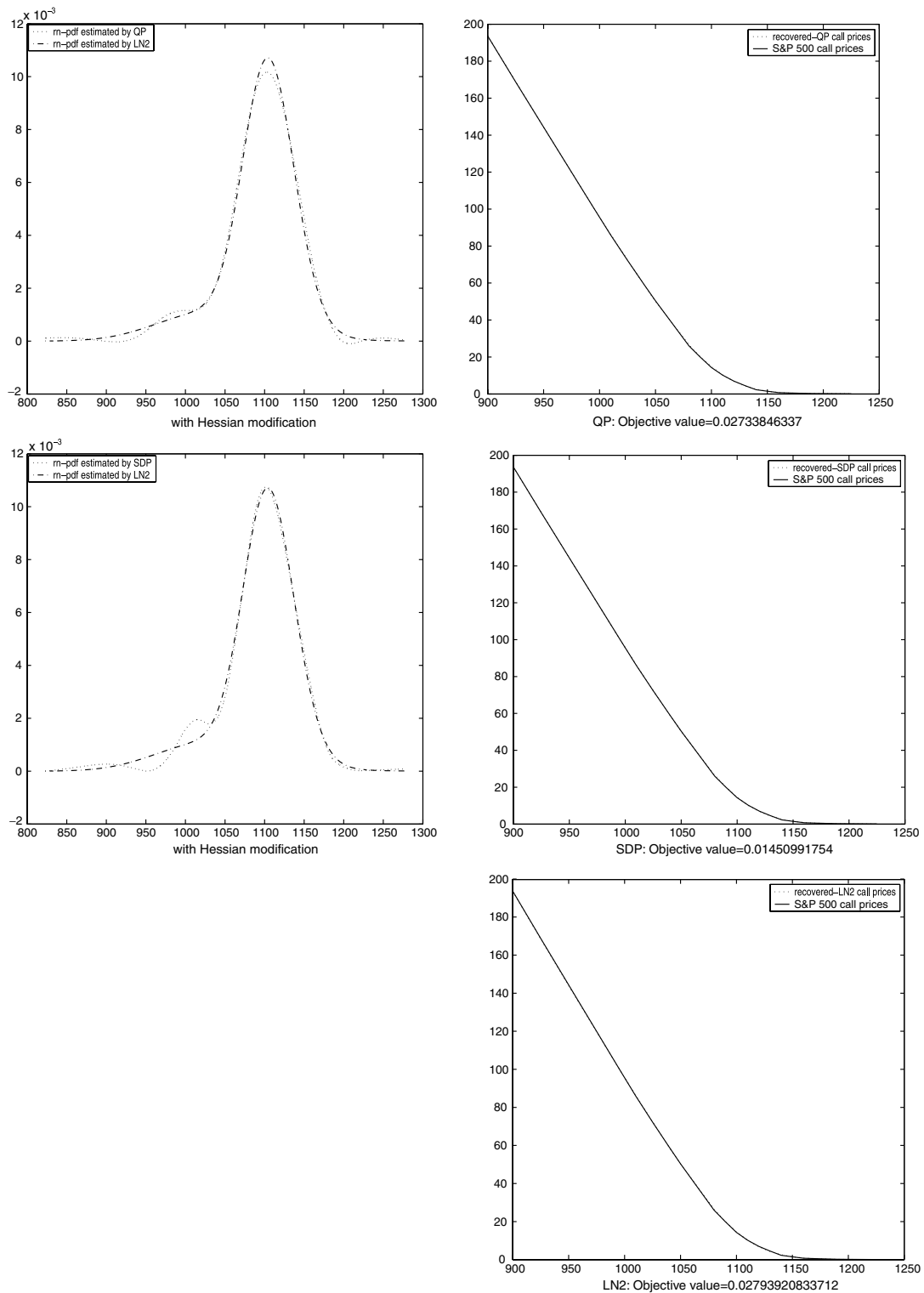


Fig. 3. Recovered probability density functions from S&P 500 Index data using QP, SDP, and LN2 approaches (left plots). Fitted recovered expected prices for these approaches (right plots). Data set 2.

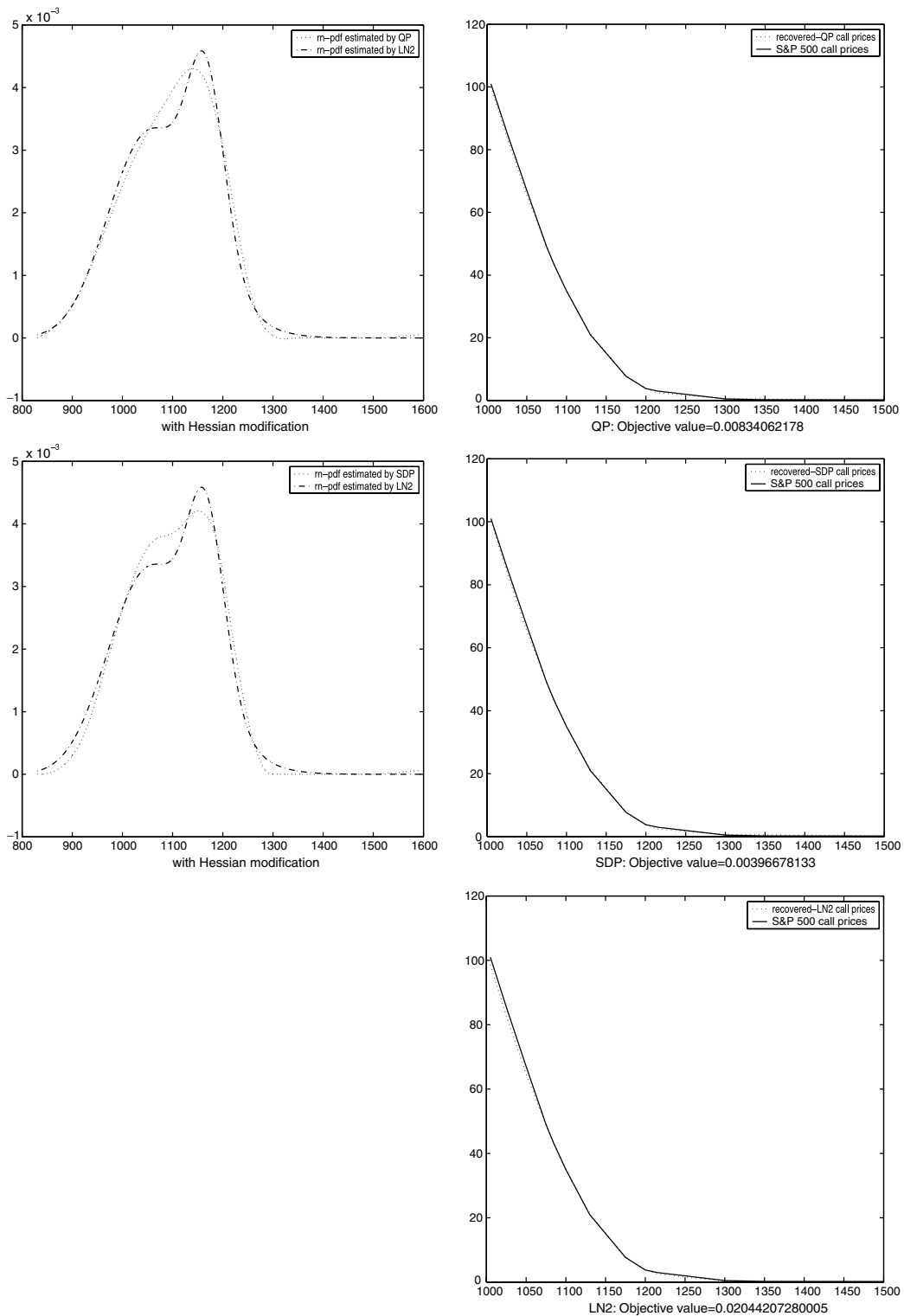


Fig. 4. Recovered probability density functions from S&P 500 Index data using QP, SDP, and LN2 approaches (left plots). Fitted recovered expected prices for these approaches (right plots). Data set 3.



$$\sum_{K \in \mathcal{C}} \theta_K [C_K - C_K(y)]^2 + \sum_{K \in \mathcal{P}} \mu_K [P_K - P_K(y)]^2$$

(and by modifying the objective function of problem (18) similarly). Here  $\theta_K$  is the ratio between the trading volume for the option  $C_K$  and the trading volume for all options:

$$\theta_K = \frac{\text{trading volume for } C_K}{\text{trading volume for all call options}}.$$

The weight  $\mu_K$  is defined similarly for put options. Note that options with zero volume have a weight equal to zero.

Table 2  
Residuals for the price option adjustments (S&P 500 Index data)

	LN2	QP	SDP
S&P500 data set 1	0.05549748812	0.04408317006	0.03413501095
S&P500 data set 2	0.02793921868	0.02733846337	0.01450991754
S&P500 data set 3	0.02142219670	0.00834062178	0.00396678133

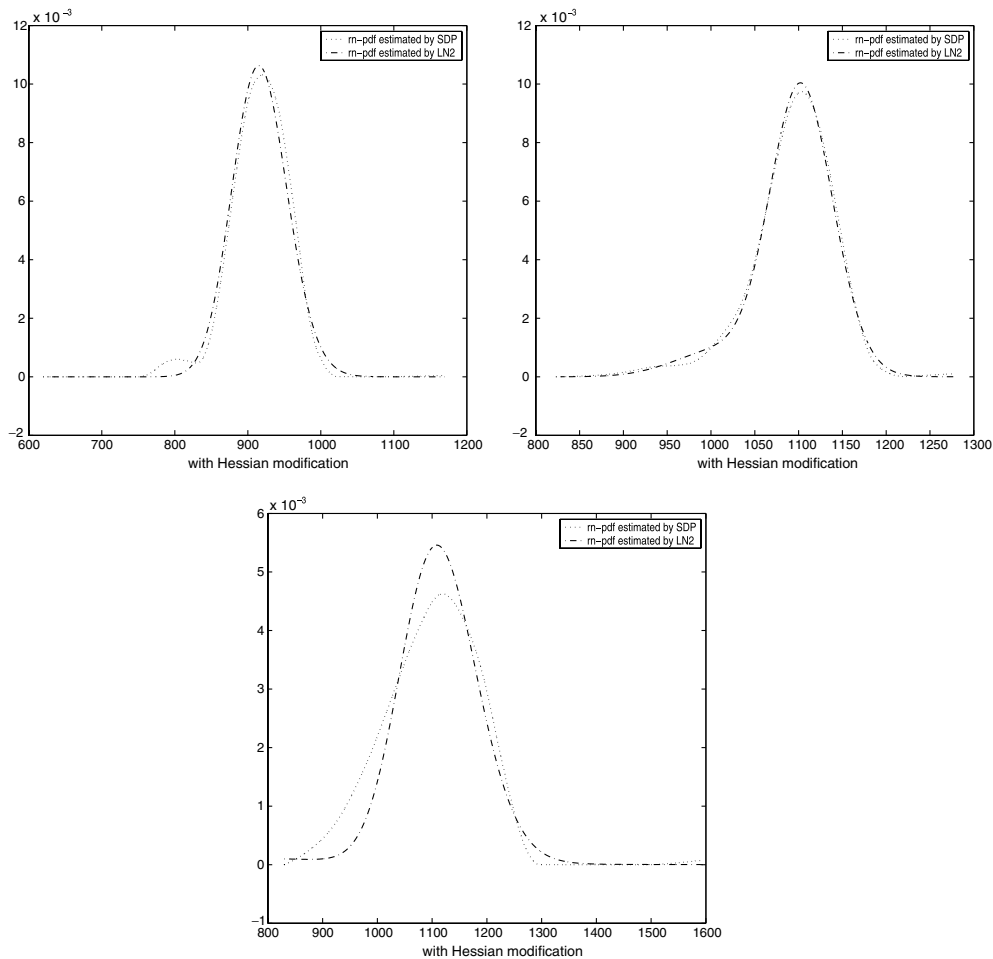


Fig. 5. Recovered probability density functions from S&P 500 Index data using arbitrage data, respectively for data sets 1, 2, and 3.

### 5.3.2. Results

The results are presented for the three data sets mentioned before, in a manner similar to the Black–Scholes case. In the first data set (Fig. 2) the original number of calls and puts was 40 each. After eliminating arbitrage opportunities we reduced the problem dimension to 24 calls for which we considered 32 knots. In the second data set (Fig. 3) the original number of calls and puts was 38 each. After eliminating arbitrage opportunities we reduced the problem dimension to 23 calls for which we considered 28 knots. Finally, in the third data set (Fig. 4) the original number of calls and puts was 29 each. After eliminating arbitrage opportunities we reduced the problem dimension to 14 calls for which we considered 29 knots. In the QP and SDP approaches for the S&P 500 data, our experiments have shown that the spline knots must be chosen close to the strike prices and that the choice of the location of the spline nodes is more sensible than it was in the Black–Scholes simulated data.

The Hessian modification has been done by adding  $\xi I$  to the scaled Hessian matrix, choosing the reference value  $\xi = |\lambda_{\min}|$  for QP and  $\xi = |\lambda_{\min}|10^3$  for SDP, both adjusted for the Black–Scholes data.

For data sets 1 and 2 the recovered probability density functions obtained by QP and SDP approaches presented bimodality in the lower tail, a feature which has not been observed in the LN2 approach. When we consider the expected prices of the call options computed using the recovered risk-neutral pdf, we see that they adjust well to the S&P500 prices (see Table 2). The fitting for the QP and SDP approaches is always better than in the LN2 case.

We have observed that the pdf estimated using the QP model and the Hessian modification assumes small negative values ( $10^{-5}$ ) at one of the tails of the distribution, roughly between 700 and 735 (Fig. 2), between 1200 and 1223 (Fig. 3), and between 830 and 870 (Fig. 4). As prescribed, the semidefinite optimization model corrects this behavior and obtains a nonnegative pdf estimate. Moreover, we see from the figures in the table that the SDP fitting is always better than the QP one.

In order to measure the impact of data containing arbitrage opportunities on our approach and the shape of the pdf generated, we repeated our tests with market data. This time, for each data set, we use the mean of bid and ask prices as call prices and did not eliminate prices that may contain arbitrage opportunities. As we can see from Fig. 5, the shape of the pdfs is similar to the shape of the previous ones (without arbitrage) for the SDP approach, indicating some robustness in our approach. For data set 3, we found some discrepancies in the LN2 approach, where the pdf estimated from data containing arbitrage differs significantly from the pdf estimated with “clean” data.

## 6. Concluding remarks

We have developed and tested a new way of recovering the risk-neutral probability density function (pdf) of an underlying asset from its corresponding option prices. Our approach is nonparametric and uses cubic splines. The core inversion problem is a quadratic programming (QP) problem with a convex objective function and linear equality constraints.

To guarantee the nonnegativity of the inverted risk-neutral pdf we followed two alternatives. In the first one we kept the QP structure of the core problem, adding linear inequalities that reflect only the nonnegativity of this pdf at the spline nodes. The second one extends the nonnegativity requirement to the entire domain of the recovered pdf by imposing appropriate semidefinite constraints. In the examples tested, we observed that the QP approach is less sensitive to scaling than the semidefinite programming (SDP) approach. While the simpler QP approach is generally sufficient to recover an appropriate risk-neutral pdf both with simulated and market data, there are instances where the solution of the more difficult SDP model is necessary to obtain a nonnegative pdf estimate.

We plan to investigate the numerical estimation of the volatility based on the knowledge of the previously estimated risk-neutral pdf. Another topic of future research is to consider uncertainty in the data and to study the robust solution of the corresponding uncertain QP and SDP problems.

## Appendix

We recall in this appendix the definition of spline functions. Consider a function  $f : [a, b] \rightarrow \mathbb{R}$  to be estimated by using its values  $f(x_s)$  given on a set of points  $x_s$ ,  $s = 1, \dots, n_s + 1$ , called nodes or knots. It is assumed that  $x_1 = a$  and  $x_{n_s+1} = b$ .

**Definition 3.** A *spline function*, or *spline*, is a *piecewise* polynomial approximation  $S(x)$  to the function  $f$  such that the approximation agrees with  $f$  on each node  $x_s$ , i.e.,  $S(x_s) = f(x_s)$ ,  $s = 1, \dots, n_s + 1$ . A spline on  $[a, b]$  is of order  $q$  if: (i) its first  $q - 1$  derivatives exist on each interior knot; (ii) the highest degree for the polynomials defining the spline function is  $q$ .

The graph of a spline function  $S$  contains the data points  $(x_s, f(x_s))$ . A spline function  $S$  is always continuous on  $[a, b]$ . A cubic spline uses cubic polynomials of the form  $f_s(x) = \alpha_s x^3 + \beta_s x^2 + \gamma_s x + \delta_s$  to estimate the function in each interval  $[x_s, x_{s+1}]$  for  $s = 1, \dots, n_s$ . A cubic spline has order +3 and can therefore be constructed in such a way that it has second-order derivatives at each node. For  $n_s + 1$  knots  $(x_1, \dots, x_{n_s+1})$  there are  $n_s$  intervals and, therefore,  $4n_s$  unknown constants to evaluate. To determine these  $4n_s$  parameters we use the following conditions:

$$f_s(x_s) = f(x_s), \quad s = 1, \dots, n_s, \quad \text{and} \quad f_{n_s}(x_{n_s+1}) = f(x_{n_s+1}), \quad (19)$$

$$f_{s-1}(x_s) = f_s(x_s), \quad s = 2, \dots, n_s, \quad (20)$$

$$f'_{s-1}(x_s) = f'_s(x_s), \quad s = 2, \dots, n_s, \quad (21)$$

$$f''_{s-1}(x_s) = f''_s(x_s), \quad s = 2, \dots, n_s, \quad (22)$$

$$f'_1(x_1) = 0 \quad \text{and} \quad f''_{n_s}(x_{n_s+1}) = 0. \quad (23)$$

The last condition leads to a so-called *natural* spline.

## References

- [1] K. Abadir, M. Rockinger, Density functionals with an option-pricing application, *Econometric Theory* 19 (2003) 778–811.
- [2] Y. Ait-Sahalia, A. Lo, Nonparametric estimation of state-price densities implicit in financial asset prices, *The Journal of Finance* 53 (1998) 499–547.
- [3] I. Anagnou, M. Bedendo, S. Hodges, R. Thompkins, The relation between implied and realised probability density functions, Technical Report, Financial Options Research Centre, University of Warwick, 2002.
- [4] S. Aparicio, S. Hodges, Implied risk-neutral distribution: A comparison of estimation methods, Technical Report, Financial Options Research Centre, University of Warwick, 1998.
- [5] D. Bertsimas, I. Popescu, On the relation between option and stock prices: A convex programming approach, *Operations Research* 50 (2002) 358–374.
- [6] B. Bhara, Implied risk-neutral probability density functions from options prices: Theory and application, Technical Report, Bank of England, 1997.
- [7] F. Black, M. Scholes, The pricing of options and corporate liabilities, *Journal of Political Economy* 81 (1973) 637–659.
- [8] R. Bliss, N. Panigirtzoglou, Testing the stability of implied probability density functions, *Journal of Banking and Finance* 26 (2002) 381–422.
- [9] O. Bondarenko, Estimation of risk-neutral densities using positive convolution approximation, *Journal of Econometrics* 116 (2003) 85–112.
- [10] D. Breeden, R. Litzenberger, Prices of state-contingent claims implicit in options prices, *Journal of Business* 51 (1978) 621–651.
- [11] B. Brunner, R. Hafner, Arbitrage-free estimation of the risk-neutral density from implied volatility smile, *The Journal of Computational Finance* 7 (2003) 75–106.
- [12] P.W. Buchen, M. Kelly, The maximum entropy distribution of an asset inferred from option prices, *Journal of Financial and Quantitative Analysis* 31 (1996) 143–159.
- [13] J. Campa, K. Chang, R. Reider, Implied exchange rate distributions: Evidence from OTC option markets, *Journal of International Money and Finance* 17 (1998) 117–160.
- [14] J. Cox, S. Ross, The valuation of options for alternative stochastic processes, *Journal of Financial Economics* 3 (1976) 145–166.
- [15] D. Duffie, *Dynamic Asset Pricing Theory*, Princeton University Press, Princeton, NJ, 1992.
- [16] D.Y. Dupont, Extracting risk-neutral probability distributions from options prices using trading volume as a filter, *Economic Series* 104, Institute for Advanced Studies, Vienna, 2001.
- [17] S. Herzel, Arbitrage opportunities on derivatives: A linear programming approach, *Dynamics of Continuous, Discrete and Impulsive Systems, Series B: Applications and Algorithms* 12 (2005) 589–606.
- [18] J.C. Jackwerth, Option-implied risk-neutral distributions and implied binomial trees: A literature review, *The Journal of Derivatives* 7 (1999) 66–82.
- [19] J.C. Jackwerth, M. Rubinstein, Recovering probability distributions from options prices, *The Journal of Finance* 51 (1996) 1611–1631.
- [20] M.S. Lobo, L. Vandenberghe, S. Boyd, H. Lebret, Applications of second-order cone programming, *Linear Algebra and Its Applications* 284 (1998) 193–228.

- [21] S. Mayhew, On estimating the risk-neutral probability distribution implied by option prices, Technical Report, Purdue University, 1995.
- [22] W.R. Mellick, C. Thomas, Recovering an asset's implied pdf from options prices: An application to crude oil during the Gulf Crisis, *Journal of Financial and Quantitative Analysis* 32 (1997) 91–115.
- [23] M. Rubinstein, Implied binomial trees, *The Journal of Finance* 49 (1994) 771–818.
- [24] M. Rubinstein, Edgeworth binomial trees, *The Journal of Derivatives* 5 (1998) 20–27.
- [25] D. Shimko, Bounds of probability, *Risk* 6 (1993) 33–37.
- [26] M. Stutzer, A simple nonparametric approach to derivative security valuation, *The Journal of Finance* 51 (1996) 1633–1652.
- [27] R.H. Tütüncü, K.C. Toh, M.J. Todd, Solving semidefinite-quadratic-linear programs using SDPT3, *Mathematical Programming* 95 (2003) 189–217.